

### Multivariate Statistical Methods for Engineering and Management (MEMEG, 1<sup>st</sup> Semester, 2021/2022)

Handed out on 10 of November, 2021.

To be handed back on 4 of December, 2021.

#### Group 3:

Mariana Silva	90685
Manuel Reis	87062
Patrícia Calhau	90687
Miguel Gouveia	90686
Miguel Simões	78386
Isaac Felizardo	90656

Consider the **LifeCycleSavings** data set, available in R.

1. Make an exploratory analysis, using plots and summary statistics (e. g. mean, covariance, generalized/total variance and Mahalanobis distances, to describe the data).
2. One researcher has rudimentary knowledge about multiple linear regression analysis and wants your help to find a way to explain the variable **sr** with some predictor variables.
  - (a) Make a preliminary analysis of the data and discuss what you have learned from this analysis.
  - (b) Fit a regression model to the dataset.
  - (c) Test for significance of the regression. Discuss the results in terms of the p-value. Compare the test results with the coefficient of multiple determination. Is there any evidence that a subset of the original variables should be excluded from the model? Proceed in order to find the best subset of regressors.
  - (d) Check model adequacy, investigate possible influential/leverage observations and outliers.
  - (e) Calculate 90% confidence interval (CI) on the mean responses for countries **Sweden** and **Portugal**. For the same values of the regressors, and the same confidence level, calculate the prediction interval (PI). Compare and discuss the obtained results.

**About the report:**

- The report should not exceed 20 pages (including Annexes).
- Do not forget to include: introduction, the dataset in study, objectives of the study, methodology used, decisions, conclusions and references.
- The R code and the report must be send to me: `irodrig@math.tecnico.ulisboa.pt` and also a print copy.